Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential

Carl F. Craver†‡

Hodgkin and Huxley’s model of the action potential is an apparent dream case of covering-law explanation in biology. The model includes laws of physics and chemistry that, coupled with details about antecedent and background conditions, can be used to derive features of the action potential. Hodgkin and Huxley insist that their model is not an explanation. This suggests either that subsuming a phenomenon under physical laws is insufficient to explain it or that Hodgkin and Huxley were wrong. I defend Hodgkin and Huxley against Weber’s heteronomy thesis and argue that explanations are descriptions of mechanisms.

1. Introduction. There is perhaps no better example of covering-law explanation in all of biology than Hodgkin and Huxley’s (HH) model of the action potential. The HH model treats the neuronal membrane as an electrical circuit involving a capacitor and three parallel series of batteries and variable resistors. One series represents the flux of sodium (Na⁺) across the neuronal membrane, another represents the flux of potassium (K⁺), and the last represents the leakage current for other charged particles. The batteries in each series represent the driving force on Na⁺, K⁺, and leakage currents, respectively, as calculated using the Nernst equation. The variable resistors represent changes in conductivity across the membrane. If one describes each of the component currents using Ohm’s law,

†To contact the author, please write to: Department of Philosophy, Philosophy-Neuroscience-Psychology Program, Washington University in St. Louis, One Brookings Drive, Wilson Hall, St. Louis, MO 63130; e-mail: ccraver@artsci.wustl.edu.
‡Thanks to Anna Alexandrova, Jim Bogen, Lindley Darden, Ken Schaffner, and Marcel Weber for comments and discussion.

1. Bogen (2008, in this issue) provides reasons to question whether the total current equation is a derived law (in the sense of Hempel [1965]). For present purposes, I grant Weber (and the advocate of covering-law explanations) that Ohm’s law and the Nernst equation are laws in the traditional sense.
one can then calculate the total current crossing the membrane using the total current equation:

$$I_{\text{total}} = C dV_m/dt + G_K n^4 (V_m - E_K) + G_{Na} m^3 h (V_m - E_{Na}) + G_C (V_m - E_C).$$

The capacitive current (represented as the first addend in the sum) depends on the capacitance of the membrane ($C$) and the change of voltage ($V_m$) over time ($t$). The $Na^+$ and $K^+$ currents are modeled using Ohm’s law ($I = V/R$). $G_{Na}$ and $G_K$ represent the maximum conductance of the membrane to each ion, which is the inverse of the minimum resistance to that ionic current. The difference between the membrane voltage and the equilibrium potential for the item ($V_m - E$) takes the place $V$ in Ohm’s law. The coefficients $n^4$ and $m^3 h$ are used to fit the curve for the conductance changes to experimental results obtained with a voltage clamp (which allows one to detect current flow and conductance changes at fixed membrane voltages). The total current equation can be used to predict numerous electrophysiological phenomena in the squid giant axon and in neurons generally. If it suffices to explain a phenomenon that one merely shows that the phenomenon is entailed by laws of nature together with a specification of initial and boundary conditions, then the HH model is a sufficient explanation.

This conclusion conflicts with Hodgkin and Huxley’s assessment:

The agreement (between our equations describing conductance changes and the voltage clamp data) must not be taken as evidence that our equations are anything more than an empirical description of the time-course of the changes in permeability to sodium and potassium. An equally satisfactory description of the voltage clamp data could no doubt have been achieved with equations of very different form, which would probably have been equally successful in predicting the electrical behaviour of the membrane. It was pointed out in Part II of this paper that certain features of our equations were capable of a physical interpretation, but the success of the equations is no evidence in favour of the mechanism of permeability change that we tentatively had in mind when formulating them. (1952, 541; italics added)

Hodgkin makes similar remarks in his Nobel Prize lecture, and many other eminent electrophysiologists of the day echo this thought (e.g., Cole 1992; Hille 1992). Even if Hodgkin and Huxley are feigning modesty, they offer arguments, and those arguments help to illustrate the virtues of mechanistic explanation in neuroscience and other special sciences.

Marcel Weber (2005) uses Hodgkin and Huxley’s achievement as an example of the “explanatory heteronomy of experimental biology.” According to Weber, all of the crucial explanatory burden in experimental
biology (including electrophysiology) is and must be borne by laws of physics and chemistry. Ohm’s law, the Nernst equation, and Coulomb’s law do the explanatory work of the model; the rest specifies mere background conditions. In contrast, I argue that crucial explanatory work is borne by distinctively biological facts—facts about conductance changes across the membrane. I argue more broadly that covering-law explanations are too weak to capture the norms of mechanistic explanation. Whereas Bechtel and Abrahamson (2005) and Machamer, Darden, and Craver (2000) argue against the covering-law model on the grounds that there are no laws in biology, I propose that we grant for now that the Hodgkin and Huxley equations make essential and unproblematic appeal to laws of physics and chemistry. My objection to the covering-law model and Weber’s heteronomy thesis, rather, is that they are too weak to capture the distinctions between description and explanation, between explanation sketches and (more) complete explanations, and between how-possibly and how-actually explanations.

2. Causal/Mechanical Explanation. Adequate explanations in neuroscience work not by showing that an explanandum phenomenon was to be expected on the basis of the laws of nature, but rather by showing how the explanandum phenomenon is situated in the causal structure of the world. Wesley Salmon (1984) argued for his view by considering cases of etiological explanation, in which an effect is explained by its cause. Here, I extend these arguments to cases of constitutive explanation, in which an explanandum phenomenon (such as the action potential) is explained through the organized activities of its component parts (such as ions and ion channels). Derivational models of explanation, and so of reduction, are too weak to capture the norms of constitutive explanation.

Hodgkin and Huxley wanted to explain the action potential by describing its mechanism. As currently understood, that mechanism begins when the axon hillock (the part of the cell connecting the cell body to the axon) depolarizes, meaning that the membrane potential rises above its resting value of roughly $-70 \text{ mV}$. This depolarization opens voltage-sensitive $\text{Na}^+$ channels in the membrane. As these channels open, $\text{Na}^+$ rushes into the cell from the relatively $\text{Na}^+$-rich extracellular space, raising the membrane voltage past zero and toward the $\text{Na}^+$ equilibrium potential (peaking at roughly $40 \text{ mV}$). As the membrane voltage crosses a threshold, voltage-sensitive $\text{K}^+$ channels open, allowing intracellular $\text{K}^+$ to diffuse out of the cell. The membrane voltage drops. At about the same time, $\text{Na}^+$ channels inactivate, blocking the outward flow of $\text{Na}^+$. As the $\text{K}^+$ and $\text{Na}^+$ channels gradually reassume their resting conformations and the $\text{Na}^+/$$\text{K}^+$ pump redistributes the ions, the membrane gradually returns to rest. (See further details about channels below.)
This is a mechanism in the sense that it is an organized collection of all and only the component entities and activities relevant to the explanandum phenomenon. The entities are the parts in the mechanism, such as the ion channels, the ions, and the Na\(^+\)/K\(^+\) pump. The activities are what these entities do, such as diffusing, attracting, repelling, blocking, and opening. These entities and activities are organized spatially and temporally such that together they make up the action potential. The axon hillock is very dense in Na\(^+\) channels (location). The ion channel has a configuration that opens into a pore (shape), and the pore is large enough for ions to pass (size). Charges are distributed through the ion-channel proteins (sequence), and they are situated in the membrane so as to respond to the effects of the changing membrane potential (orientation). Different channels activate at different times (temporal sequence), and they stay open on average for well-defined temporal intervals (duration). To provide a mechanistic explanation, one shows how the different features of the phenomenon depend upon the organization of the entities and activities that compose the mechanism (see Craver 2007; cf. Machamer et al. 2000).

3. Hodgkin and Huxley’s Modesty. If Hodgkin and Huxley are right, one needs to know the above mechanism (and a good deal more) to explain the action potential. Hodgkin and Huxley argue against the explanatory import of their mathematical model on three related grounds.

First, their model provides nothing more than an empirical description of the time course of the permeability changes in the membrane. Kenneth Cole likewise claims that the HH model “summarized in one neat tidy little package the many thousands of experiments done previous to 1952, and most subsequent ones” (1992, 151). Figure 1 shows an action potential superimposed on curves representing the conductance changes for Na\(^+\) (left) and K\(^+\) (right) as determined on the basis of voltage clamp experiments. Hodgkin and Huxley’s key mathematical achievement was to find equations for the values of \(n\), \(m\), and \(h\) in the total current equation that fit the conductance curves. The current equation for K\(^+\) involves the expression \(n^4\). The current equation for Na\(^+\) involves the expression \(m^3h\). The values of \(n\), \(m\), and \(h\) change as a function of voltage and time (see Weber 2008, in this issue, for fuller discussion).

Hodgkin and Huxley note that their “carrier model,” according to which ions are actively transported across the membrane, cannot explain the data obtained from the voltage clamp. However, they conclude that

---

2. I am not making the absurd claim that no explanation can be represented in mathematical form. My point is that the equation, absent causal interpretation, does not explain the action potential.
We soon realized that the carrier model could not be made to fit certain results, for example the nearly linear instantaneous current voltage relationship, and that it had to be replaced by some kind of voltage-dependent gate. As soon as we began to think about molecular mechanisms it became clear that the electrical data would by themselves yield only very general information about the class of system likely to be involved. So we settled for the more pedestrian aim of finding a simple set of mathematical equations which might plausibly represent the movement of electrically charged gating particles. (Hodgkin 1992)

The HH model (particularly the “activation molecules” that Hodgkin and Huxley imagined in building their mathematical models for conductance changes) is in this respect more analogous to Ptolemy’s planetary models—which neither involve nor imply any commitment to the existence of the epicycles, deferents, and equants from which they are constructed—than it is to Newton’s gravitational model of planetary motion, which Newton presented to show how and why the planets move as they do. The model embodies no commitments about the mechanisms that change the membrane conductance, allow the ionic currents to flow, and coordinate them so that the action potential has its characteristic shape. In the HH model, commitments about underlying mechanisms are replaced
by mathematical constructs (the equations for $n$, $m$, and $h$) and hypothetical entities (activation molecules) that save the phenomena much as Ptolemy’s geometrical constructions save the apparent motion of the planets through the night sky. They do not show how the membrane actually changes its permeability. As Hodgkin and Huxley write, “Details of the mechanism will probably not be settled for some time” (1952, 504).

Hodgkin and Huxley wanted to replace knowledge of the shape of the conductance change curves with knowledge of the causal mechanisms that explain why the curves have the shape that they do. Nothing in the equations for conductance change captures the difference between mere temporal sequences and causal relationships or, more generally, between mere correlates and causal relations. The equation itself does not specify whether the rise in $\text{Na}^+$ conductance causes the rise in $\text{K}^+$ conductance, nor does it give any indication of why the curves have the slope that they do or why they inactivate at a given threshold. It is true that the equations represent membrane conductance (permeability) as dependent upon voltage. But mathematical dependencies cannot be equated with causal or explanatory dependency relations. The equations must be supplemented by a causal interpretation: one might, for example, agree by convention that the effect variable is represented on the left, and the cause variables are represented on the right, or one might add “these are not mere mathematical relationships among variables but descriptions of causal relationships in which this variable is a cause and this other is an effect,” but the point is that one will have to specify which variables represent causes and which represent effects, and one will have to specify which of the myriad mathematical relationships contained in the equations stand for causal relations and which stand for mere correlations. In other words, nothing in the HH models for the conductance changes specifies which mathematical relations are accidents and which are explanatory laws (i.e., descriptions of causal dependencies).

To be sure, Hodgkin and Huxley knew a good deal more about action potentials than they include explicitly in their mathematical model. Adding this detail helps to turn a mechanism sketch, a model that has black boxes, into a more complete description of a mechanism. Hodgkin and Huxley knew that the action potential is produced by changes in membrane conductance to $\text{Na}^+$ and $\text{K}^+$. They knew that ions flux across the membrane toward their equilibrium potentials. They knew that this flux of ions constitutes a transmembrane current. This background knowledge about the mechanism does provide a partial explanation for how neurons

---

3. It turns out that aspects of the equations for conductance change do correspond to features of ion channels, but, as Hodgkin and Huxley would have noted, it is not a perfect fit and, at any rate, it is merely fortuitous.
generate action potentials because it reveals some of the components of the mechanism, some of their properties, and some of their activities. The total current equation supplements this background knowledge with explicit temporal constraints on the mechanism. The equations include variables that represent important components in the explanation. And the equations provide powerful evidence that a mechanism built from those components could possibly explain the action potential. The equations, supplemented with a diagram of the electrical circuit in a membrane and supplemented with details about how membranes and ion channels work, carry considerable explanatory weight. The equations without such interpretation—an interpretation that is difficult for those who know much more than Hodgkin and Huxley did about the mechanism of the action potential to forget—do not constitute the explanation. Hodgkin and Huxley’s first argument is exactly right.

Skip now to their third. They insist that they have “no evidence” in favor of the mechanism that they “tentatively had in mind” when formulating their equations. According to that hypothesized mechanism, the membrane’s permeability to Na\(^+\) is regulated by the position of four particles: three “activation molecules” that move from the outside of the membrane to sites on the inside, and one “inactivation molecule” that can block either the activation molecules or the flow of Na\(^+\) through the membrane. The expression \(m_3 h\) in the total current equation can then be interpreted as the joint probability that all three activation molecules are in their open state (\(m\) is the probability that any one molecule has moved) and that no inactivation molecule is bound (\(h\)). When Hodgkin and Huxley say that they have no evidence for their hypothesized mechanism, they are referring to this vaguely specified activation molecule mechanism. The choice of a different strategy for building the equation (e.g., using a single variable, or three rather than two) would suggest a different physical interpretation. At most, this simple model of the activation and inactivation of Na\(^+\) channels provides a how-possibly sketch of the action-potential mechanism—a description of the sort of mechanism that could account for the action potential but is not, in fact, how neurons generate action potentials.

Return then to their second argument: that the curves for conductance changes could have been achieved with equations of a different form. In 1963, Huxley reflected that “it was clear that the formulation we had used was not the only one that might have fitted the voltage clamp results adequately” (Huxley 1963, 61). Indefinitely many equations could be used to predict the conductance changes and so the action potential’s time course. And these different mathematical equations might be given any number of biological interpretations, such as the activation molecule model. Hodgkin and Huxley had no reason to privilege this one how-
possibly model above the others as a how-plausibly or how-actually model. To explain the action potential required that they know further details about the molecular mechanisms underlying the conductance changes. Such details restrict the space of possible mechanisms to a subspace of plausible mechanisms that will presumably be narrowed further with time.

In the 1950s and 1960s, talk of ‘channels’ involved an imaginative leap that few researchers were willing to accept. Hille describes his effort to construct a space of possible mechanisms:

In the next decade, Clay and Armstrong and I began our independent research. In our first papers, we brought a clear list of ‘molecular’ assumptions to the table. They included the following ideas: ions are passing through aqueous pores that we called channels, ion channels are proteins, the channels for Na⁺ and K⁺ are different, they have swinging gates that open and close them, we can study their architecture by using electric currents to measure gating, permeation and block, and channel blockers are molecules that enter the pores and physically plug them. (Hille et al. 1999, 1106)

The term ‘channel’, however, was viewed by many as merely a filler term for an activity or mechanism to be named later:

From 1965 to 1973, such ideas were debated annually at the meetings of the Biophysical Society. There, prominent scientists would routinely rise to request that anyone who chose to use the word “channel” avow first that it bears absolutely no mechanistic implications! It is probably fair to say that people thought the discussion about molecular mechanisms was premature. In 1969, when I had drafted a summary review of these ideas, Kenneth Cole, the dean of American biophysics, wrote to me: “I’m . . . worried you may be pushing some of your channel arguments pretty far.” (Hille et al. 1999, 1106)

The idea of activation molecules (let alone pores or gates) was a useful fiction—a how-possibly model—for Hodgkin and Huxley. It helped them to model the action potential, but it cannot be interpreted in terms of details about neuronal membranes. Hille and his colleagues move beyond this useful fiction by positing a set of how-possibly models, assessing them on experimental and theoretical grounds, and so constraining the space of possible mechanisms. To leap to the end of the story, it is now well known that conductance changes across the membrane are effected by conformation changes in ion-specific channels through the cell membrane. Biochemists have isolated these channels, sequenced their constituents, and learned a great deal about how they activate and inactivate during an action potential. It is this wealth of detail about how channels regulate
the timing of the conductance changes, as described by the HH equations, that explains the temporal course of the action potential.

So Hodgkin and Huxley are right to acknowledge the explanatory limitations of mere mathematical models. First, there is a difference between knowing a mere temporal sequence and knowing the causal relations among components. The deductive consequences of the model are the same whatever one’s interpretation of the causal structure that the model represents. Second, there is a difference between phenomenal models, which merely characterize the explanandum phenomenon, and mechanistic models, which explain why the phenomenon has its characteristic features. The equations governing the values of \( n \), \( m \), and \( h \) are merely phenomenal models.\(^4\) They characterize how specific ion conductances change with voltage and time, but they do not explain why they change as they do, when they do. Without an account of the underlying mechanisms of the conductance change, the buck of accounting for the various temporal features of the action potential is merely passed on to some conductance-changing process, we know not what. Even if one includes the bulk of background knowledge available to scientists at the time and well into the 1970s, they at best could sketch the mechanism, not describe it completely. Finally, Hodgkin and Huxley admirably recognize the distinction between a how-possibly model (such as their activation model) and a how-actually (or even how-plausibly) explanation. They knew that activation particles were purely conjectural, and they would not accept conjectures as explanations.

4. **Weber’s Heteronomy Thesis.** Consider now Weber’s claim that the HH model exemplifies his heteronomy thesis for experimental biology:

It is a physico-chemical law that ultimately explains why ions move across membranes when action potentials spread. Most of the residual parts of the mechanistic model described above basically describe conditions under which this ion transport occurs. Thus, the entire model could be viewed as an explanation that uses covering laws from physical chemistry in combination with a description of the initial and boundary conditions that obtain in neural cell membranes. It is these covering laws that have explanatory force; the rest of the mechanistic model basically states how the physico-chemical theory should be applied (Weber 2005, 26; see also 28–29)

The heteronomy thesis makes two claims: (1) that physico-chemical laws are necessary for explanation in biology and (2) that genuinely biological

---

terms have no explanatory import besides designating the system to which the covering laws apply.\(^5\)

Weber’s argument for claim (1) is independent of the HH example. He argues that the laws of biology lack the necessity of physico-chemical laws because they are historically contingent. Others claim that the generalizations fail as laws because they are mechanistically fragile, hedged with ceteris paribus clauses, and limited in scope. Elsewhere, I argue that generalizations with these features can explain nonetheless, so long as those generalizations continue to hold under well-controlled experimental manipulations (cf. Woodward 2003). In this issue, Weber explicitly relinquishes claim (1) on similar grounds in order to challenge my claim that the HH model fails as a complete explanation of the action potential.

Turn now to claim (2). It is historically inaccurate to claim that Hodgkin and Huxley worked out the basic mechanisms of the action potential in the 1950s and, especially, to suggest that they had any knowledge of voltage-gated channels (see Weber 2005, 21). Hodgkin and Huxley knew that their electrophysiological methods could not decide among competing mechanisms of conductance change. That is why they claim that their model is incomplete. They knew, I argue, that it is insufficient to explain the action potential merely to show that its form can be derived from laws of physics and chemistry. It is insufficient, first, because the laws for the changes in membrane conductance (laws that fail to satisfy Weber’s earlier demand for physical necessity in genuine explanations) are essential for explaining why the action potential has its characteristic shape. It is insufficient, second, because it fails to specify the biological details about how the membrane conductance changes. The currently accepted explanation appeals to the behavior of voltage-sensitive ion channels, behavior that is also historically contingent, limited in scope, and mechanistically fragile. Descriptions of channel structure and function are not mere “descriptions of initial and boundary conditions”; they are crucial features of the mechanism for the action potential.\(^6\) Any putative explanation that leaves them out is missing something crucial.

5. Explanation in Neuroscience. The point of developing a model of explanation in neuroscience and other special sciences is to illuminate the

5. Weber does not specify his procedure for sorting the information in explanations into those components that do explanatory work and those that are merely background and boundary conditions. Nor does he specify a model of explanation. Without specifying such a procedure and providing such a model, it is difficult to see how Weber can mount a full defense of claims (1) and (2).

6. I switch between speaking of the permeability of the membrane to ions and conductance changes to current more or less interchangeably.
difference between good and bad explanations. The primary advantage of inferential accounts of explanation, such as the covering-law model, is that they express clear criteria of adequacy and completeness for explanations. It is well known that these models face serious challenges in capturing the norms of causal (unilevel) explanations, as exemplified in classic examples such as the flagpole and the shadow, the hexed salt, and the barometer and the storm (Salmon 1984). Yet the discussion of inter-level explanations retains implicit and vestigial commitments to an inferential model of explanation.

The HH model illustrates the limitations of inferential accounts of explanation by showing that argumentative subsumption under general regularities (including the ability to derive a phenomenon from highly general equations grounded in laws of physics) is insufficient for explanation. It is common in the philosophy of biology to criticize inferential models (such as the covering-law model) because there are no laws of biology. As noted above, this is one primary motivation for Weber’s heteronomy thesis: the explanatory force in biological explanations must come from physics and chemistry because it cannot come from the laws of biology. Hodgkin and Huxley’s arguments, as I have reconstructed them, attack instead the sufficiency of inferential models: the norms of explanation embodied in inferential models are too permissive to distinguish good neuroscientific explanations from bad ones. Such models of explanation face the challenge of distinguishing phenomenal models from models of mechanisms, mechanism sketches from complete descriptions of mechanisms, and how-possibly from how-actually explanations. A more appropriate regulative ideal demands (as Hodgkin and Huxley appear to demand) that explanations accurately describe the relevant portion of the causal structure of the world (see Salmon 1984; Craver 2007). The norms of constitutive explanation follow from a commitment on the part of researchers to describe mechanisms accurately and completely.

To say that the HH model is insufficient as an explanation of the action potential is not to deny the model’s fundamental importance. The equation serves many functions in electrophysiology: as a data summary, a characterization of the phenomenon to be explained, a set of temporal constraints on the mechanism, a predictive tool, and a heuristic guide to future theorizing (see Bogen 2005). Of particular importance is the epistemic role of Hodgkin and Huxley’s mathematical demonstration. Their model is the basis for a potent argument that the observed changes in membrane conductance during an action potential could account for diverse features of the action potential in physiologically relevant conditions. The body of research preceding the HH model, authored by B. Katz, R. D. Keynes, H. J. Curtis, and Cole, can be described as a gradual attempt to add piecemeal constraints on the space of possible mecha-
nisms: to determine that current flows across the membrane, to determine which ions carry the current, to characterize the time course and magnitude of the action potential, to measure intra- and extracellular ionic concentrations. The fundamental achievement of Hodgkin and Huxley’s model was to show that a mechanism involving variable conductances to ionic currents could reproduce features of the action potential within known constraints of how those conductances change, on the ionic composition of the internal and external milieu, and so on. One should not mistake this epistemic achievement, an argument in favor of a given mechanistic explanation, for the explanation itself. Arguments are not explanations. Mechanisms explain action potentials.

REFERENCES
Copyright of Philosophy of Science is the property of Philosophy of Science Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.